Development of a highly optimized procedure for the discovery of RNA G-quadruplexes by combining several strategies

Marc-Antoine Turcotte, François Bolduc, Anaïs Vannutelli, Jérémie Mitteaux, David Monchaud, Jean-Pierre Perreault

Development of a highly optimized procedure for the discovery of RNA G-quadruplexes by combining several strategies

Marc-Antoine Turcotte [a], François Bolduc [a], Anaïs Vannutelli [a], Jérémie Mitteaux [b], David Monchaud [b], Jean-Pierre Perreault [a, *]

[a] Department of Biochemistry and Functional Genomics, Pavillon de Recherche Appliquée sur le Cancer, Université de Sherbrooke, Sherbrooke, Quebec, J1E 4K8, Canada
[b] Institut de Chimie Moléculaire de l'Université de Bourgogne, CNRS, UMR 6302, Dijon, 21078, France

A B S T R A C T

RNA G-quadruplexes (rG4s) are non-canonical secondary structures that are formed by the self-association of guanine quartets and that are stabilized by monovalent cations (e.g. potassium). rG4s are key elements in several post-transcriptional regulation mechanisms, including both messenger RNA (mRNA) and microRNA processing, mRNA transport and translation, to name but a few examples. Over the past few years, multiple high-throughput approaches have been developed in order to identify rG4s, including bioinformatic prediction, *in vitro* assays and affinity capture experiments coupled to RNA sequencing. Each individual approach had its limits, and thus yielded only a fraction of the potential rG4 that are further confirmed (i.e., there is a significant level of false positive). This report aims to benefit from the strengths of several existing approaches to identify rG4s with a high potential of being folded in cells. Briefly, rG4s were pulled-down from cell lysates using the biotinylated biomimetic G4 ligand BioTASQ and the sequences thus isolated were then identified by RNA sequencing. Then, a novel bioinformatic pipeline that included DESeq2 to identify rG4 enriched transcripts, MACS2 to identify rG4 peaks, rG4-seq to increase rG4 formation probability and G4RNA Screener to detect putative rG4s was performed. This workflow uncovers new rG4 candidates whose rG4-folding was then confirmed *in vitro* using an array of established biophysical methods. Clearly, this workflow led to the identification of novel rG4s in a highly specific and reliable manner.

1. Introduction

An RNA G-quadruplex (RNA G4 or rG4) structure is a secondary structure that is found in guanine rich RNA. It is formed when the guanines self-assemble into quartets which then self-stack in the presence of monovalent cations (typically potassium (K^+)). The minimal canonical motif is « 5′-$G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$-3' », where N can be any ribonucleotide [1]. However, multiple examples of non-canonical G4s present in RNA, as well as in DNA, are found in the literature where bulges, long-loops, missing guanines and triads are all part of the G4 [2]. These non-canonical G4s are usually rejected after classical bioinformatic sequence based analysis, which implies that substantially more G4s could be identified than was initially predicted [3]. Moreover, since pattern matching search methodology does not consider neighboring sequences that can prevent formation of rG4 by favoring alternative Watson-Crick-based structures, its use led to many false-positive predictions. Second generation of G4 prediction algorithms, based on an alternative to pattern matching search, permit the prediction of significantly more potential rG4 motifs (see below).

rG4s are of interest as genetic targets because of their implication in multiple post-transcriptional regulation mechanisms such as microRNA (miRNA) maturation, messenger RNA (mRNA) polyadenylation, alternative splicing, translation and protein sequestration, to name just a few of them [4–8]. Given this growing interest, several high-throughput techniques have been developed in order to identify new G4s on both the genome and the transcriptome scales [9]. First, the folding probability of a putative G-quadruplex-forming sequence (or pG4) can be determined exclusively on the basis of its nucleotide sequence. Among the reported tools, some of them rely solely on the presence of a G4 motif, on the

* Corresponding author.
E-mail address: jean-pierre.perreault@usherbrooke.ca (J.-P. Perreault).

List of abbreviations

CDS	Coding sequence
CD	Circular dichroism
G4	G-quadruplex
K^+	Potassium
Li^+	Lithium
MACS2	Model-based Analysis of ChIP-Seq 2
mRNA	Messenger RNA
miRNA	Micro RNA
ncRNA	Non-coding RNA
NMM	N-methyl mesoporphyrin
scRNA	Small-cytoplasmic RNA
snRNA	Small-nucleolar RNA
PDS	Pyridostatin
pG4	Putative G-quadruplex-forming sequence
rG4	RNA G-quadruplex
RTS	Reverse transcription stalling
TDS	Thermal difference spectra
UTR	Untranslated regions

calculation of consecutive guanines as compared to cytosines and on neural networks developed with confirmed positive and negative G4 sequences [10–16]. Neural network techniques are particularly interesting since the scores are not human-based [14]. Secondly, purified nucleic acids can be incubated under different conditions and the presence of putative rG4s can then be detected by various techniques, including RT-qPCR and sequencing. The well-known G4-seq and rG4-seq techniques belong to this class [17,18]. These techniques rely on either the polymerase's (G4-seq) or the reverse transcriptase's (rG4-seq) stalling in order to detect G4 sites *via* differential analyses performed under Lithium (Li^+)- versus K^+-rich conditions, and in either the absence or the presence of the G4 binding ligand pyridostatin (PDS). *In vitro*, properly folded G4s can also be identified *via* their precipitation using G4-specific antibodies, such as hf2 and BG4, followed by either RT-qPCR (hf2 for DNA G4s, BG4 for RNA G4s) or sequencing (BG4 for DNA G4s) [19–22]. Thirdly, higher-order nucleic acid structures can be fixed live prior to being extracted from the cells, pulled down by G4s-specific molecular tools and identified by either RT-qPCR or sequencing. Examples include the immunoprecipitation of DNA G4s with BG4 (G4 ChIP-seq), the precipitation of DNA G4s using the engineered protein G4P (G4P ChIP-seq) and the chemo-precipitation of both DNA and RNA G4s with BioTASQ (G4DP-seq and G4RP-seq) [23–27]. BioTASQ is a biomimetic G4 ligand that interacts with both DNA and RNA G4s, and its biotin handle permits the isolation of the resulting G4/ligand complex *via* streptavidin-based precipitation [28]. The main advantage of the G4RP-seq technique is the capacity of the BioTASQ ligand to precipitated RNA from fixed cells, ensuring no change in G4 fold during the sample preparation. It has been used to isolate and identify both DNA and RNA G4s, both *in vitro* and *in vivo*.

Each of these approaches has been used with success to identify novel G4/rG4 structures. That said, in each case when looking at the rG4s' folding one by one, there was a relatively large number of false-positive candidates identified at the validation step. This step is time-demanding and can be expensive. In order to overcome this hurdle, it was decided to combine some of these techniques with the aim of identifying rG4s with high specificity. More specifically, G4 isolation and sequencing (G4RP) and *in silico* analysis (rG4-seq and G4RNA Screener) were implemented to uncover new rG4s (Fig. 1).

2. Material and methods

2.1. G-quadruplex RNA precipitation (G4RP)

In order to identify new rG4s, G4RP was performed as described previously, with minor modifications [24,25]. Briefly, a confluent 15 cm Petri dish of U87MG cells was washed once at room temperature with sterile Phosphate Buffered Saline (PBS) and then incubated for 5 min with 25 mL of 1% formaldehyde/PBS on a rocker at room temperature. The reaction was then quenched by the addition of 3.75 mL of 0.125 M glycine, and the sample was then incubated on a rocker for 5 min at room temperature. The solution was removed, and the cells were washed once with PBS and then scraped with 2 mL of PBS. The cells were then harvested by centrifugation at 400×*g* for 3 min at room temperature. The supernatant was removed, and another wash was performed with 2 mL of PBS. The cells were then resuspended in 1 mL of sterile G4RP buffer (25 mM Tris-HCl pH 7.4, 150 mM KCl, 5 mM EDTA, 1 mM DTT, 0.5% NP-40 and 2000 U/mL of homemade RNase inhibitor). The sample was then sonicated on ice at 20% amplitude for 180 cycles of 1 s ON − 3 s OFF using a Branson Digital Sonifier 450. The resulting solution was centrifuged at 17 000×*g* for 10 min at 4 °C. The supernatant was separated into 2 tubes, 450 µL for the BioTASQ incubation and 50 µL for the input. The BioTASQ ligand was then added to the cellular extract (50 µL of 1 mM BioTASQ) and the sample was incubated overnight on a rotor at 4 °C. The input was kept at 4 °C. The next morning, 40 µg of PBS-washed Streptavidin-magnetic beads (Promega) were added to the 500 µL solution and the entire mix was incubated for 2 h at 4 °C on a rotor. The beads were then separated using a magnet for 2 min. They were then washed 3 times for 5min with G4RP buffer and once for 2 min with PBS, all at room temperature. The washed beads were then resuspended in 100 µL of PBS, while 50 µL of PBS was added to the input. Both solutions were reverse cross-linked for 1.5 h at 70 °C. The RNA was then isolated using QIAzol according to the manufacturer's protocol (Qiagen), ethanol precipitated and dissolved in sterile water. A DNAse incubation was then performed on both samples using the Illustra RNA Spin Mini RNA kit according to the manufacturer's protocol (General Electric). Ribosomal RNAs were removed using the NEBNext rRNA Depletion kit Human/Mouse/Rat according to the manufacturer's protocol (New England Biolabs).

2.2. Libraries and sequencing

Libraries and sequencing were performed at the Laboratoire de Génomique Fonctionnelle de Université de Sherbrooke. The libraries were performed using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) as recommended by the manufacturer. The triplicate was sequenced on a NEXT-Seq 500 (Illumina) sequencer at 2 × 45 bp. A minimum of 35 million reads were detected in each sample.

2.3. Reverse transcription and digital PCR

RNAs were quantified by digital RT-PCR at the Laboratoire de Génomique Fonctionnel de Université de Sherbrooke. For each reverse transcription, 100 ng of RNAs were used. The primers used for the digital PCR are listed in Supp. File S5. For the biotin condition, the BioTASQ was replaced with biotin in the G4RP protocol.

2.4. Bioinformatic analysis

The sequenced reads were analyzed using FastQC, trimmed using Trimmomatic and mapped on the reference genome GRCh38.p13 (GENCODE) using STAR [29–32]. Read distributions in

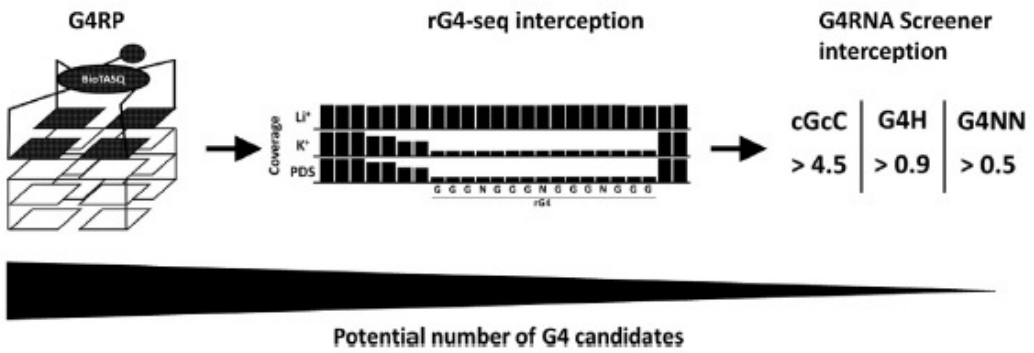


Fig. 1. Overview of the protocol used to identify rG4s with high fidelity. rG4s were detected by peak-calling using the sequencing results from the BioTASQ ligand condition as compared to those obtained from the input condition. In order to compensate for the possible non-specificity of the BioTASQ ligand, filters were added. First, the peaks were compared with the rG4s identified using a high-throughput technique, rG4-seq, under either the K^+ or the PDS conditions. Secondly, the hits were compared with potential G4s identified using G4RNA Screener.

the 5′UTR, the CDS and the 3′UTR, and the G4 densities, were analyzed according to a protocol reported previously [3]. In order to perform the gene differential expression analysis between the BioTASQ and the input condition, the reads mapped by STAR were counted for exons and were annotated using FeatureCount [33]. The normalization and differential expressions were performed using DESeq2 in R [34]. Further analyses were performed using Excel. Genes with a base mean of 0, a negative $\log_2$FC and an adjusted p-value greater than 0.05 were removed from the analysis. In order to perform the peak calling, the reads mapped by STAR were extended to generate fragments using BamCoverage with the following options: -of bigwig; -bs 1; –effectiveGenomeSize 2913022398; and, -e 125 [35]. Then, the bigwig files were converted with bigWigToBedGraph into BedGraph files [36]. Next, one replicate at a time, MACS2 bdgcmp was used with the option « -m ppois » to deduct the noise of each input from the BioTASQ condition. Peaks were found using MACS2 bdgpeakcall with the following options: –cutoff 5; –min-length 50; and, –max-gap 0 [37]. The peaks with scores over 200 were kept for further analysis. The peaks present in each triplicate were found using bedtools Intersect intervals [36]. Reads and peaks were intersected with 5′UTR, CDS and 3′UTR regions from the GRCh38.p13 gff file. The GC content for the G4RP-seq MACS2 peaks was found by extracting the peaks sequences with the tool Extract genomic DNA in Galaxy and then calculating the different nucleotide content with Excel. Bedtools was also used for the determination of the intersection of the triplicate, the rG4-seq data and the pG4 [3,18]. G4 density were found in the Vannutelli manuscript [3]. The selection of random genes to evaluated random density was repeated 100 times. Data were visualized on IGV [38]. The sequencing data have been deposited into NCBI's BioProject and are accessible at PRJNA833051.

2.5. Fluorescence assays

In order to evaluate G4 formation, the RNAs were dissolved at a concentration of 2 μM in a solution containing 100 mM of either KCl or LiCl and 20 mM of Li-cacodylate (pH 7.5). The samples were then heated at 70 °C for 5 min and slowly cooled down at room temperature over a period of 1 h. N-methyl mesoporphyrin (NMM) was then added to a final concentration of 4 μM (2:1 ratio). The fluorescence was then evaluated as described previously [39]. All experiments were performed in duplicate.

2.6. Circular dichroism (CD) and thermal denaturation spectra (TDS)

In order to evaluate the RNAs' structures, the RNAs were dissolved at a concentration of 4 μM in a solution containing 100 mM of either KCl or LiCl and 20 mM of Li-cacodylate (pH 7.5). The samples were then heated at 70 °C for 5 min and slowly cooled down to room temperature over a period of 1 h. The experiments were performed in a quartz cell with a path length of 1 mm on a Jasco J-810 spectropolarimeter. Data were recorded from 350 to 210 nm, at a rate of 50 nm per min with a 2 s response time, 0.1 nm pitch and 1 nm bandwidth [40]. The CD experiments were performed at 20 °C, and the TDS at both 20 and 90 °C. All CD spectrums were smoothed. TDS spectrums were obtained by subtracting the 90 °C absorbance spectrum from the 20 °C absorbance spectrum and normalizing the highest value to 1 [41].

3. Results

3.1. Protein coding RNAs are enriched with the G4RP protocol

In order to identify novel rG4s, the G4RP protocol was applied to the glioblastoma cell line U87MG, a cell line that is often used in both neuroscience and immuno-oncology research [25,42]. Briefly, the cells were fixed with formaldehyde before being lysed and the nucleic acids fragmented by sonication. These extracts were then incubated with BioTASQ and the resulting nucleic acids/BioTASQ complexes were pulled-down using streptavidin-coated magnetic beads. After separation from the beads and reversal of the cross-linking by heat treatment, the RNA fragments were then isolated by phenol-based extraction. After two additional treatments (DNAse and ribodepletion), the purified RNAs were then ready to be sequenced and analyzed.

The read counts obtained from the sequencing data derived from the various conditions were similar, except for the second of the three replicates, in which more reads were found to be present in the BioTASQ condition as compared to the input one (Fig. S1 and see the further discussion below). This could be explained by the fact that a different batch of BioTASQ has been used for each replicate to reduce the variability that one batch could have. The efficiency of the protocol was verified *via* RT-ddPCR amplifications, which showed a notable pull-down of the G4-containing NRAS, TERF2 and VEGFA RNAs, but not of a non-G4-containing control RNA (U6 RNA; Fig. S2A), in the BioTASQ condition compared to the use of biotin molecules, as was reported in the original G4RP experiments (Fig. S2B) [28].

The transcriptomic regions enriched by BioTASQ were further scrutinized. As pG4s are known to be enriched in both the 5′- and the 3′-untranslated regions (5′- and 3′UTRs, respectively) [3,10,18], the ability of G4RPs to exhibit a similar trend was evaluated. The reads obtained for each BioTASQ triplicate (plus the input) were mapped in either the annotated 5′UTR, the coding sequence (CDS)

or in the 3′UTR regions. The read distribution was then evaluated (Fig. 2A–C). As expected, an enrichment of reads in the BioTASQ condition was observed in the 5′UTR. However, reads were also enriched in the CDS and were relatively absent from the 3′UTR, which points towards an unusual pattern of G4 distribution in the U87MG cells.

Subsequently, a differential expression analysis was performed with DESeq2 in order to identify transcripts that are enriched in the presence of BioTASQ as compared to the input condition. A principal component analysis was performed on the DESeq2 data to evaluate the variance between each replicate and each condition. The plot showed, as expected, two distinct clusters: one for the input and one for the BioTASQ condition (Fig. S3). The transcripts with a positive $\log_2$ fold-change (i.e. those enriched under the BioTASQ condition) and an adjusted p-value of at least 0.05 were kept for further analysis (Supp. File S1). In total, 3254 transcripts were found to be enriched (Fig. S4). Their significance increased when both the number of fragments mapping on each transcript (base mean counts) and the fold-change increased. Among them, 3024 were classified as protein-coding RNA (93%), 63 as non-coding RNA (ncRNA, 2%), 37 as pseudogenes (1%), 2 as small cytoplasmic RNA (scRNA, 0.06%), 2 as small nucleolar RNA (snRNA, 0.06%) and 126 that remained as unclassified RNAs (4%). In order to evaluate the type of transcript enriched by BioTASQ, the 20 most abundant candidates with the highest fold-changes and a base mean of at least 500 were analyzed (Fig. 2D). Most of these transcripts (80%) were classified as protein-coding RNAs, as was reported previously for MCF-7 cells [24]. Only RN7SL3 (scRNA), RN7SL1 (snRNA), LOC285908 (pseudogene) and BCYRN1 (ncRNA) were classified otherwise.

Finally, the pG4 densities were evaluated in both the 100 transcripts with the highest fold change and in those with the 100 lowest fold change using the exon sequences and a stringent bioinformatic approach [3]. As expected, these results showed that the top 100 BioTASQ pulled-down genes had a higher pG4 density compared to the bottom 100 and a random set of 100 genes (i.e., calculated 100 times) (Fig. S5). These differences were also observed in the abundant candidates, which is similar to the results previously reported in the original BioTASQ experiment. Although those results highlighted the presence of predicted G4s in the transcripts pulled down by BioTASQ, their positions were still unknown. It was therefore decided to further investigate the enrichment at local positions on the genes instead of the enriched genes themselves.

3.2. Specific peaks were retrieved with BioTASQ

The G4RP-seq dataset was further analyzed using a peak-calling approach so as to identify regions that are specifically enriched in the presence of BioTASQ. For each replicate, the input background was subtracted from the BioTASQ coverage. The Model-based Analysis of ChIP-Seq 2 technique (MACS2) [37] was used and identified between 11 978 and 85 043 peaks for each replicate (Fig. 3A). When an arbitrary MACS2 cut-off score of 100 was applied, the number of peaks was reduced by roughly 50%; however, many of the peaks still presented a low G4 enrichment profile between the BioTASQ/input conditions. The cut-off was then raised to 200, which further decreased the number of candidate peaks by 35%, resulting in the identification of between 1634 and 19 185 peaks for each replicate. Among them, 797 peaks were found to be

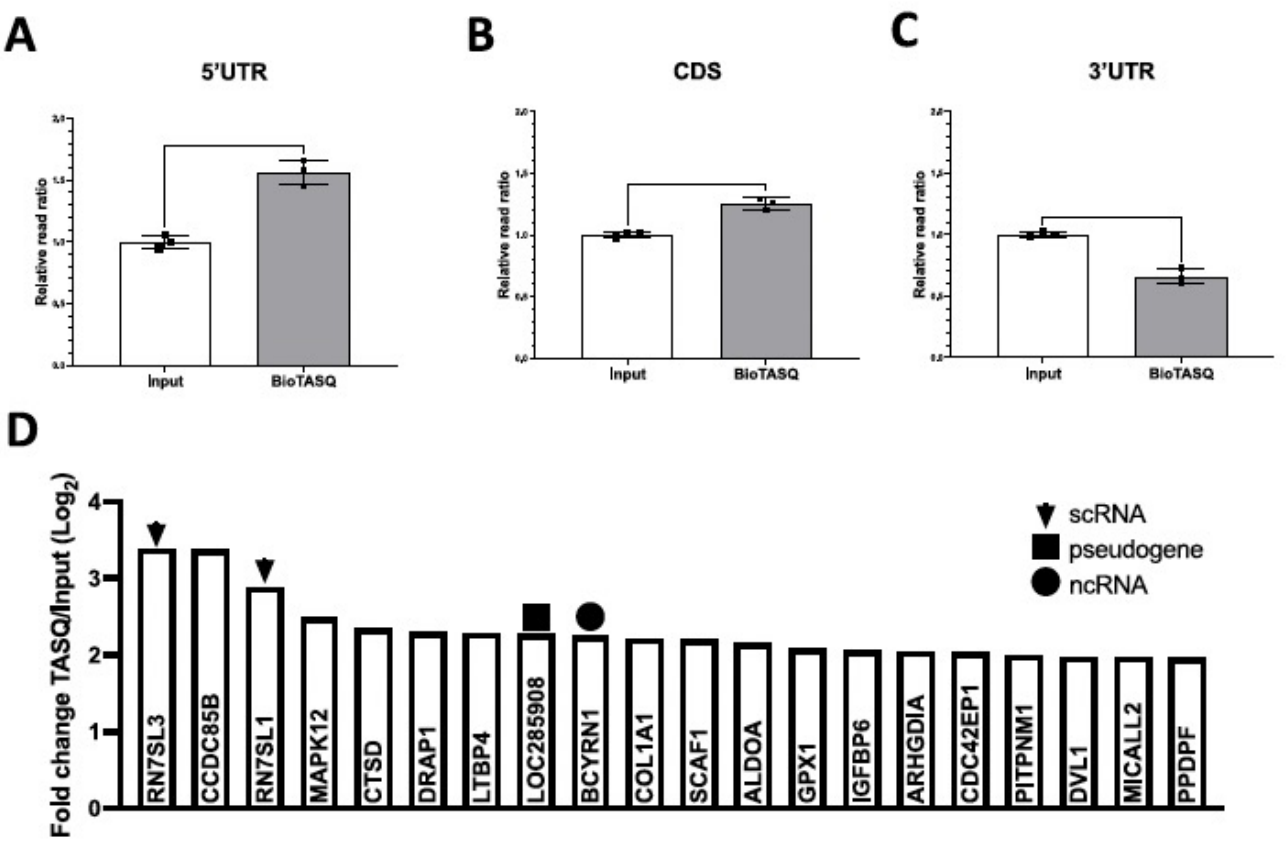


Fig. 2. Read distributions and genes enriched in the presence of the BioTASQ ligand. Analysis of the read distribution between the input and the BioTASQ conditions in the: A) 5′UTR; B) CDS; and, C) 3′UTR. The Y-axis represents the number of reads normalized to the number of reads in the input. The error bars represent the error between the three replicates. The significance of the different conditions was evaluated using a *t*-test (* represents a p-value of less than 0.05). D) The 20 most abundant candidates identified as being the most enriched in the presence of the BioTASQ ligand are presented in histogram form. The Y-axis represents their fold change (Log_2). The genes annotated as small cytoplasmic RNA (scRNA) are marked with a triangle, that annotated as pseudogene is marked with a square and the gene annotated as a non-coding (ncRNA) is marked with a circle. All the unmarked genes are annotated as protein-coding mRNAs.

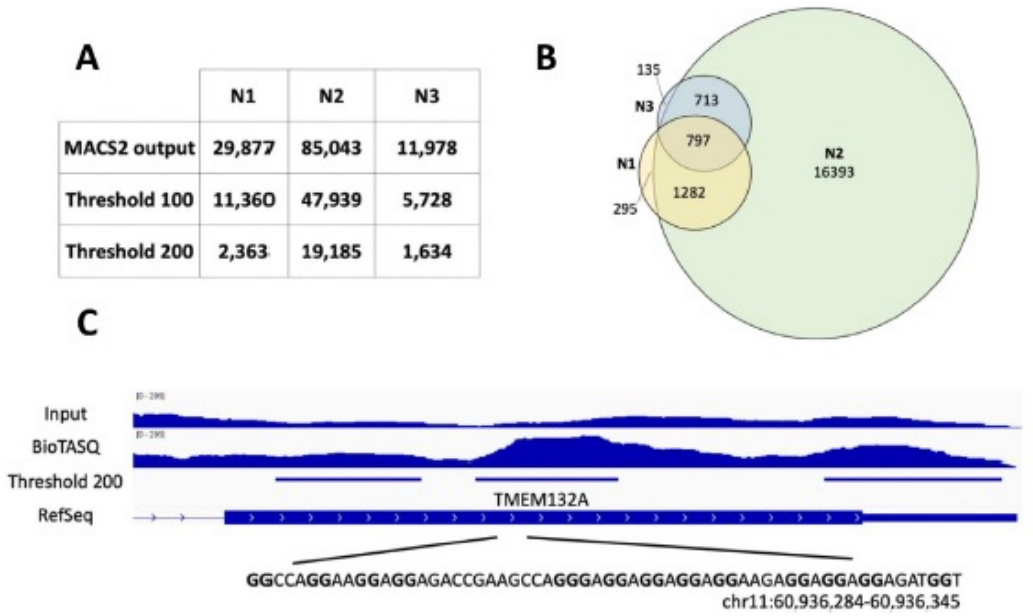


Fig. 3. Identification of new potential rG4s by peak-calling. A) Number of peaks identified by MACS2, and the impact of the different thresholds on each triplicate (no threshold applied, a 100 or a 200 threshold). B) Venn graph of the peaks identified in each replicate. C) Representative example in the TMEM132A gene of the coverage observed when the BioTASQ condition is compared to that of the input. The peak found with MACS2 is shown in the third row. The genomic coordinates and sequence of a possible G-quadruplex with two quartets in this peak are shown.

common to all replicates and were kept for further analysis (Fig. 3B and Supp. File S2). These peaks were highly guanine- and cytosine-rich (66%, further discussed below) on their transcript strand (and consequently adenine- and thymine-poor, 34%), thus being different from the 41% G/C-content normally observed in the human genome (Fig. S6) [43]. As an example, MACS2 identified the RNA corresponding to the transmembrane protein TMEM132 (Fig. 3C), for which the enrichment between the BioTASQ/input conditions can clearly be seen. A potential G4-forming sequence within this peak is also highlighted in the figure.

In terms of peaks enrichment, G4RP peaks were equally found in the 5′UTR, CDS and 3′UTR (32, 32 and 36%, respectively). Compared to the BioTASQ reads' distribution (10, 59 and 31%), the G4RP peaks are less distributed in the CDS and more in the 5′UTR. Finally, when the number of peaks was normalized to the size of the different regions, there was a clear enrichment of peaks in the 5′UTR (23 for the 5′UTR, 4 for the CDS and 3 peaks/Mb for the 3′UTR). Therefore, these data go in the same way as for the reads' distribution, showing an unusual G4 distribution pattern in the U87MG. All the following bioinformatic interception were also made with the cut-off of 100, but not further presented (See Fig. S7).

3.3. *Peaks refinement improves the quantity of G4-like sequences*

The 797 peaks identified were aligned with the data from the rG4-seq performed in human HeLa cells [18]. This cross-cell lines analysis allows for the identification of G4s conserved amongst cell lines, which likely increases the specificity of the method. The rG4-seq technique identified G4 sites as reverse transcription stalling (RTS) sites under either K^+-rich conditions or K^+-rich conditions in the presence of PDS and compared them to what was obtained in non-G4 promoting Li^+ conditions. Initially, rG4-seq identified 3845 RTS in the presence of K^+, and 13 423 RTS sites in the presence of both K^+ and PDS (Supp. File S3). The G4RP regions obtained possessed 43 peaks in common with the rG4-seq data obtained in the K^+ condition, and 116 with the rG4-seq data obtained in the PDS condition, representing 5% and 15% of the initial G4RP peaks (Fig. 4A, S8 and Supp. File S3). Several regions containing a previously characterized G4 were also found in the analysis. Some, such as H2AFY [44], were detected by both the G4RP and the rG4-seq in the K^+ condition (G4RP/K^+), and both VEGFA and TERC (along with H2AFY [44–46]) were detected by both the G4RP and the rG4-seq techniques in the presence of PDS (G4RP/PDS). The rG4-seq data also annotated the RTS into different categories: canonical G4s, non-canonical G4s (including long loops, bulged and 2-quartet G4s) and others. For both the G4RP/K^+ and the G4RP/PDS G4s, 41% and 22% of the identified G4s were annotated as canonical, and 59% and 75% as non-canonical, respectively. Note that 12% (K^+) and 15.3% (PDS) of the RTS in the rG4-seq data were classified as other, likely false positives, as compared to 0% (K^+) and 3% (PDS) with the G4RP/rG4-seq dataset.

In order to still further confirm the G4 nature of these transcripts, and to further bolster the specificity of the workflow, the G4RNA Screener bioinformatic tool was used [47]. More specifically, the G4RP/rG4-seq candidates were compared with the human pG4s [3] using a stringent threshold (i.e. the three scores needed to be positive in order for the candidate to be considered as being positive (Supp. File S4)). This led to 25 G4s being identified by G4RP, rG4seq/K^+ and G4RNA screener (G4RP/K^+/pG4); and to 54 G4s being identified by G4RP, rG4seq/PDS and G4RNA screener (G4RP/PDS/pG4), representing 3 and 7% of the initial G4RP peaks, respectively (Fig. 4A, S8 and Supp. File S4). The previously characterized H2AFY G4 was still detected under these conditions. For the hits from the G4RP/K^+ and the G4RP/PDS G4s, 40% and 26% of the identified G4s were annotated as canonical, and 60% and 74% as non-canonical, respectively, indicating that the bioinformatic filter does not affect the category proportions. Interestingly, 23% of the G4RP peaks were originally found in the G4RNA Screener candidates. Therefore, the list is reduced by 20 and 16% when the intercepted with the rG4-seq data are added (Fig. S8). Altogether, the combination of these three techniques resulted in the reliable detection of established rG4s and the identification of 79 new rG4 candidates.

3.4. *The potential candidates fold into G-quadruplexes*

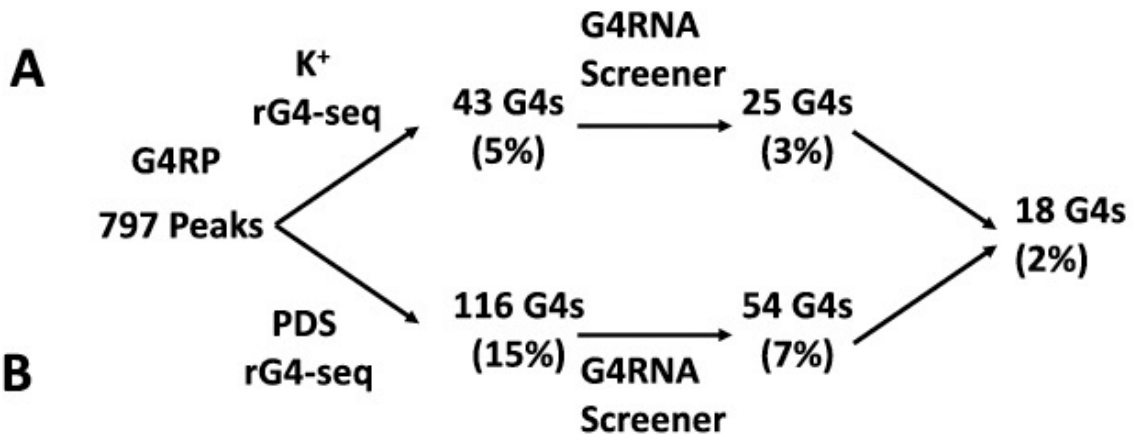
In order to validate the potential rG4 candidates, they were further analyzed using three complementary biochemical

techniques: the N-methyl mesoporphyrin (NMM) fluorescence assay, circular dichroism (CD) and thermal difference spectra (TDS) [41,48,49]. These investigations were performed with the 17 sequences that are in common to both the G4RP/K^+/pG4 and the G4RP/PDS/pG4 (H2AFY was not investigated as its folding has already been demonstrated both *in vitro* and *in cell*) (Fig. 4B) [44]. Briefly, the 17 novel RNAs were synthesized via T7 RNA polymerase-based *in vitro* transcription from the appropriate DNA oligonucleotides (Supp. File S5) [50]. They were then tested for their ability to trigger the fluorescence of NMM, looking for a different response when the sequences were incubated in the presence of either Li^+ (unfolded, weak fluorescence) or K^+ (folded, strong fluorescence). As seen in Fig. 5A and S9, most, if not all, RNAs provided a significant differential response, with the notable exception of TFE3. As an example, the fluorescent spectra of TMEM132A are shown in Fig. 5B where the NMM-G4 characteristic emission maxima at 605 nm is used for calculations.

Next, the ability of all the RNA sequences to fold into G4s was evaluated by CD. The typical CD spectra of rG4s are characterized by a negative peak at 240 nm and a positive one at 264 nm, which reflects their parallel conformation [51]. However, this technique is usually performed with short sequences in order to minimize the effects of a mixture of structures. Since the G4 sequences identified here are 60-nt long, variabilities in both the maximum and the minimum peaks should be expected. As seen in Fig. S10, all the candidates displayed a CD spectrum typical of an rG4 (including TFE3), with only a few minor variations being observed. As an example, the CD spectrum of TMEM132A is shown in Fig. 5C.

Finally, a third technique was used to validate the G4 formation, specifically the thermal difference spectra (TDS). This technique is based on the mathematical difference of the UV–vis spectra of a given sequence recorded at low (20 °C, folded) and high temperatures (90 °C, unfolded) [41]. G4s are identified by a TDS signature displaying an hypochromicity (negative peak) near 295 nm, as can be observed in the case of TMEM132A (Fig. 5D). Except for SNX14, all the candidates exhibited a TDS spectrum corresponding to G4 formation (Fig. S11).

The G4-folding was thus investigated by three biophysical methods and confirmed for 15 RNAs, while both TFE3 and SNX14 received support from only two of the three methods. Collectively, these results demonstrate the high fidelity of the pipeline with a specificity of over 85% for the identification of new rG4 candidates.



Gene Name	Sequence (5' to 3')	Position
AES	GGAUUAGCAGGGGGCCGGGACAGGGAGGUUGGGAGGGGGGACAGAGGGGAGACAGAGGCA	3'UTR
ARHGAP23	GGGAGAGGAGGUGUAGGGUGCGUAUGUCCAUGGGGGGAGGGGCUUGUGUGUGCAGUCAUU	3'UTR
ATXN2L	GGGGCUGCAGGUGGGGCAGGAUGCACGGGUUCUGGGUGGGGAGUGAGGGGUCUUGGAGGC	CDS
C19orf6	GGUCAGCAAGGAGAGGGGGUGGGGUUCCGCGGAAGGUUCUGGAGGGGUCUUGGUAGGUCU	3'UTR
EIF4H	GGAGCAAGAAUGAGCCUGCGGUUGGGAGGGAAUGGGGCGUGGGGGGUUAGAGCAGGACCA	3'UTR
GALNT2	GGAUGCGUGCGAGCUGAGGACAGGGCGGGAGGAGGGGGCACACAUGCCCCAGGGGAGCGA	3'UTR
IGFBP6	GGAAGCUGAGGGCUGUCUCAGGAGGGAGGGGCAGGAGUGCGGGGUCUACACCCCUAACUG	CDS
MAF1	GGCAACGAGCUGGACAUGGAGCUGGGGGAGGAGGAGGUGGAGGAAGAAAGCAGAAGCGGG	CDS
PPP2R4	GGCUGGCCGCCAGCUGUGGGCUGGGGUGGGGUGGGUGGGGUCGGGUGGUGGAGAGGCCUU	3'UTR
RPS9	GGAAGAAUGCCAAGAAGGGCCAGGGUGGGGCUGGGGCUGGAGACGACGAGGAGGAGGAUU	CDS
SH3GL1	GGACCCCACACUUGGGUGGGGGGGCUUAUCUGGGUGGGUGGGGAUGCCUGUUUACACUAG	3'UTR
SNX14	GGUGGGGCGGGGACGCGAGGUGUGGAUGGGGGGUCGCCUUGACCUCUGCCUCAGCCAGUA	5'UTR
SORBS3	GGCCUCCAGCUGGGUGUGGGGGGGCGGAGCAAGGCGGGGGACAGACGCAGCACCUUCUUA	3'UTR
STC2	GGGAGGAGGAGGGGAAGCGGCGAAGGAGGAAGAGGAGGAGGAGGAAGAGGGGAGCACAAA	5'UTR
TFE3	GGAGGACAUUCUGAUGGAGGAGGAGGAGGGGGUGGUGGGAGGACUGUCGGGGGGUGCCCU	CDS
TMEM132A	GGCCAGGAAGGAGGAGACCGAAGCCAGGGAGGAGGAGGAGGAAGAGGAGGAGGAGAUGGU	CDS
TMEM184B	GGGAGACCGGGAGGGGUGGCCACACAGUGGGGAGCUGGGAGGUGGGGGGAAUGGUCCCAG	3'UTR

Fig. 4. rG4 candidates retained via the procedure. A) Schematic representation of the different filters added to the G4RP results. The percentages represent the remaining number of peaks as compared to the initial number of 797. B) Names, sequences and positions of the candidates identified. Only the G4s present under both the K^+ and the PDS conditions are shown.

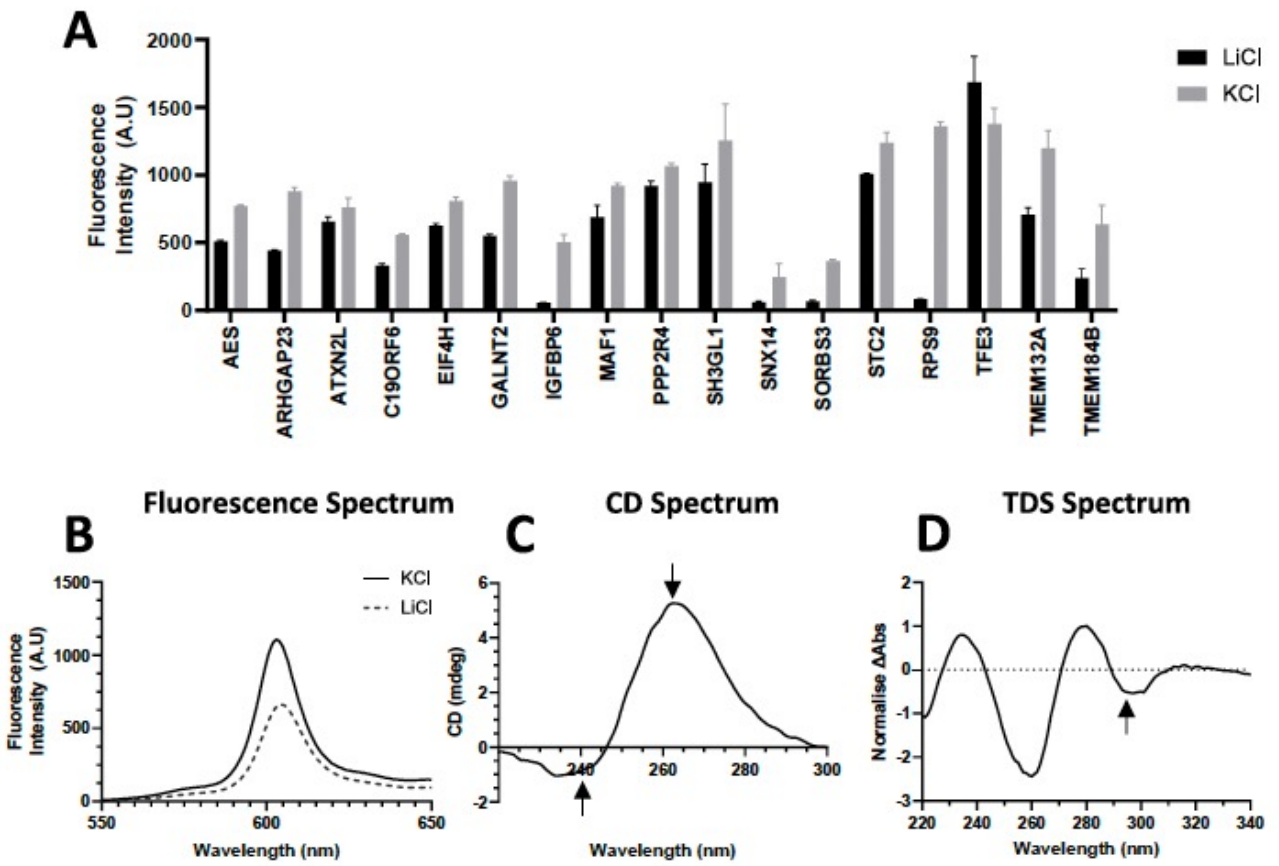


Fig. 5. *In vitro* characterization of the new RNA G4s by fluorescent assay and by both circular dichroism and thermal differential spectra. A) NMM fluorescence assays were performed in the presence of 2 μM of RNA, 4 μM of NMM and 100 mM of either KCl or LiCl at room temperature. All experiments were performed in duplicate. B) Example of the fluorescence spectra obtained for TMEM132A. C) Example of the circular dichroism spectrum obtained for TMEM132A in the presence of 4 μM of RNA and 100 mM of potassium. D) Example of the circular thermal differential spectra obtained for TMEM132A in the presence of 4 μM of RNA and 100 mM of potassium. The arrows represent the typical peaks observed for rG4s.

4. Discussion

The establishment of a rigorous procedure that leads to the identification of new rG4s with high fidelity contributes to the discovery of novel cellular mechanisms. The workflow presented here, which combines precipitation-based sequencing methods with bioinformatic and biophysical assays, allowed for the identification of 15 novel rG4 candidates that were selected based on very stringent criteria.

This workflow relies on the G4RP-seq protocol: *ca.* 3254 transcripts were isolated, the majority of which (93%) were protein-coding RNAs. Data mining (G4 peak calling using MACS2) with a proper detection threshold led to the identification of *ca.* 797 G^4 peaks (common to biological replicates) within these transcripts. These peaks were aligned with rG4-seq datasets collected in either K^+-rich conditions (referred to as "K^+"), or in K^+-rich conditions in the presence of PDS (so called "PDS"), which identified 43 common G4RP/K^+ peaks and 116 common G4RP/PDS peaks. Here, we hypothesize that many G4s (either canonical or non-canonical) found with MACS2 are not stable enough to stop the polymerase and induce a stalling site in the rG4 experiment. However, in the G4RP-seq protocol, the cell of fixed to pull-down the different G4s, which means that the stability of the G4s is not a matter. This alignment also eliminates the potential cytosines bias. These peaks were further trimmed by G4RNA screener, which led to the detection of 25 and 54 high-confidence G4 peaks, in the K^+ and the PDS conditions, respectively. The G4 structures of 17 common sequences were confirmed experimentally by NMM fluorescence titrations, CD and TDS investigations. The 17 G4s candidates characterized were enriched in the 3′UTR (7% 5′UTR, 33% CDS and 60% 3′UTR). As mentioned in the introduction, these G4s could influence the ribosomal progression (5′UTR and 3′UTR), the binding of proteins (5′UTR, CDS and 3′UTR), miRNA binding and polyadenylation (3′UTR). Of these, 15 G4s were fully confirmed, and 2 were found to be atypical. The first one, TFE3, had a higher fluorescence intensity in Li^+-rich conditions. It was hypothesized that it could fold under both K^+/Li^+ conditions but this folding is different depending on the cation and the resulting G4s could have different affinities for NMM ligand. The second one, SNX14, there was a difference in its NMM spectrums that can be observed between the potassium and lithium conditions. This is in contradiction with the absence of negative peaks at 295 nm in TDS, which points toward a non-G4 structure. This phenomenon could be explained by only a minority of sequences having folded into a G4.

The U87MG cell line was used here in order to identify new RNA G4 candidates for future studies related to neuronal diseases. G4s have been shown to be implicated in multiple neuronal diseases and dysfunctions [52–55]. However, the way the G4s are implicated in these diseases remains to be elucidated. To this end, multiple neuronal transcripts were identified with the G4RP/rG4-seq/pG4 comparison and with the OMIM database [3,18,56]. Among the 79 transcripts identified here, some are related to neurodegenerative disorders, including PSAP (Prosaposin, which is involved in Parkinson's disease) and RAB11B (Ras-related protein Rab-11B, which is involved in neurodevelopmental disorders with ataxic gait) for instance, and both of which deserve to be further investigated *in cella*.

A limitation of the study is the surprising enrichment of cytosine-rich RNAs by the BioTASQ (Fig. S6). This unprecedented observation could originate in an interaction between the

BioTASQ's guanines with these sequences' cytosines, as these cytosine-rich RNAs do not fold into i-motifs (as their DNA counterparts do in a pH-dependent manner, Figs. S12 and S13, Tables S1 and S2). The possible interaction of the cytosines of unfolded RNAs with the guanines of BioTASQ was indeed substantiated by pull-down experiments (Fig. S14 and Table S3). This observation is important as such an effect was not seen during the G4RP experiments performed in the MCF-7 cells. This provides a message of caution about a possible caveat of this technique but also emphasizes the interest of the presented pipeline, as it rectifies this effect to maximize the identification of real rG4s.

Another limitation to the use of omics for the identification of new rG4s is a bias toward the G4s present in highly expressed RNAs. As an example, RTS found in the upgraded rG4-seq protocol (rG4-seq 2.0) showed a strong correlation with the initial amount of RNA used [57]. Consequently, G4s located in RNAs that are expressed at less than 4 transcripts per million are usually not identified by this technique. The same phenomenon was observed with the peak-calling of the G4RP-seq data: MACS2 attributes a lower score to peaks with lower coverage, which indicates that using this procedure may fail to detect G4s found in lowly expressed genes.

In conclusion, this study evaluated a different strategy for the identification of new RNA G4s with high fidelity. This workflow hinges on a combination of tools including high-throughput techniques (G4RP and rG4-seq), bioinformatic predictions (G4RNA Screener) and biophysical validation (NMM fluorescence, CD and TDS). We believe that this unique combination offers the best guarantee for the identification of new RNA G4 candidates, which must be considered for subsequent *in cell* studies aimed at uncovering new genetic levers that can be manipulated for both understanding and addressing pathologies.

Author contributions

Conceptualization, T.M.A., B.F., M.D. and P.J.P; Methodology, T.M.A., B.F., M.D. and P.J.P; Formal Analysis, T.M.A, V.A and M.J.; Data curation and visualization T.M.A, V.A and M.J.; Writing – Original Draft Preparation, T.M.A., M.D. and P.J.P, Writing – Review & Editing, T.M.A., M.D. and P.J.P; All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We thank D. Bergeron, M. Avino and P.E. Jacques for their support with the bioinformatic analyses. This project was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC; 155219–17, to P.J.P.), the Centre national de la recherche scientifique (CNRS to M.D.) and the European Union (PO FEDER-FSE Bourgogne 2014/2020 programs, Grant No. BG0021532, to M.D.). T.M.A. received student fellowships from the Fonds de Recherche Québec Nature et Technologie (FRQNT) and the Canadian Institutes of Health Research (CIHR). P.J.P. holds the Research Chair of the Université de Sherbrooke in RNA Structure and Genomics and is a member of the Centre de Recherche du CHUS. The funders had no role in study design, data collection and analysis, decision to publish, nor in the preparation of the manuscript.

References

[1] S. Rouleau, R. Jodoin, J.-M. Garant, J.-P. Perreault, RNA G-quadruplexes as key motifs of the transcriptome, in: Catal. Act. Nucleic Acids, Springer International Publishing, Cham, 2017, pp. 1–20, https://doi.org/10.1007/10_2017_8.
[2] M.T. Banco, A.R. Ferré-D'Amaré, The emerging structural complexity of G-quadruplex RNAs, RNA 27 (2021) 390–402, https://doi.org/10.1261/rna.078238.120.
[3] A. Vannutelli, S. Belhamiti, J.-M. Garant, A. Ouangraoua, J.-P. Perreault, Where are G-quadruplexes located in the human transcriptome? NAR Genomics Bioinforma 2 (2020) https://doi.org/10.1093/nargab/lqaa035.
[4] P. Agarwala, S. Pandey, K. Mapa, S. Maiti, The G-quadruplex augments translation in the 5′ untranslated region of transforming growth factor β2, Biochemistry 52 (2013) 1528–1538, https://doi.org/10.1021/bi301365g.
[5] J.-D. Beaudoin, J.-P. Perreault, Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening, Nucleic Acids Res. 41 (2013) 5898–5911, https://doi.org/10.1093/nar/gkt265.
[6] K. Matsumura, Y. Kawasaki, M. Miyamoto, Y. Kamoshida, J. Nakamura, L. Negishi, S. Suda, T. Akiyama, The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration, Oncogene 36 (2017) 1191–1199, https://doi.org/10.1038/onc.2016.282.
[7] S.G. Rouleau, J.-M. Garant, F. Bolduc, M. Bisaillon, J.-P. Perreault, G-Quadruplexes influence pri-microRNA processing, RNA Biol. 15 (2017) 198–206, https://doi.org/10.1080/15476286.2017.1405211.
[8] I. Georgakopoulos-Soares, G.E. Parada, H.Y. Wong, R. Medhi, G. Furlan, R. Munita, E.A. Miska, C.K. Kwok, M. Hemberg, Alternative splicing modulation by G-quadruplexes, Nat. Commun. 13 (2022) 2404, https://doi.org/10.1038/s41467-022-30071-7.
[9] K. Lyu, E.Y.-C. Chow, X. Mou, T.-F. Chan, C.K. Kwok, RNA G-quadruplexes (rG4s): genomics and biological functions, Nucleic Acids Res. 49 (2021) 5426–5450, https://doi.org/10.1093/nar/gkab187.
[10] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic Acids Res. 33 (2005) 2908–2916, https://doi.org/10.1093/nar/gki609.
[11] O. Kikin, L. D'Antonio, P.S. Bagga, QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences, Nucleic Acids Res. 34 (2006) W676–W682, https://doi.org/10.1093/nar/gkl253.
[12] J.-D. Beaudoin, R. Jodoin, J.-P. Perreault, New scoring system to identify RNA G-quadruplex folding, Nucleic Acids Res. 42 (2014) 1209–1223, https://doi.org/10.1093/nar/gkt904.
[13] A. Bedrat, L. Lacroix, J.-L. Mergny, Re-evaluation of G-quadruplex propensity with G4Hunter, Nucleic Acids Res. 44 (2016) 1746–1759, https://doi.org/10.1093/nar/gkw006.
[14] J.-M. Garant, J.-P. Perreault, M.S. Scott, Motif independent identification of potential RNA G-quadruplexes by G4RNA screener, Bioinformatics 33 (2017) 3532–3537, https://doi.org/10.1093/bioinformatics/btx498.
[15] A.B. Sahakyan, V.S. Chambers, G. Marsico, T. Santner, M. Di Antonio, S. Balasubramanian, Machine learning model for sequence-driven DNA G-quadruplex formation, Sci. Rep. 7 (2017) 14535, https://doi.org/10.1038/s41598-017-14017-4.
[16] M. Turner, Y.M. Danino, M. Barshai, N.S. Yacovzada, Y. Cohen, T. Olender, R. Rotkopf, D. Monchaud, E. Hornstein, Y. Orenstein, rG4detector, a novel RNA G-quadruplex predictor, uncovers their impact on stress granule formation, Nucleic Acids Res. 50 (2022) 11426–11441, https://doi.org/10.1093/nar/gkac950.
[17] V.S. Chambers, G. Marsico, J.M. Boutell, M. Di Antonio, G.P. Smith, S. Balasubramanian, High-throughput sequencing of DNA G-quadruplex structures in the human genome, Nat. Biotechnol. 33 (2015) 877–881, https://doi.org/10.1038/nbt.3295.
[18] C.K. Kwok, G. Marsico, A.B. Sahakyan, V.S. Chambers, S. Balasubramanian, rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome, Nat. Methods 13 (2016) 841–844, https://doi.org/10.1038/nmeth.3965.
[19] E.Y.N. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, G-quadruplex structures are stable and detectable in human genomic DNA, Nat. Commun. 4 (2013) 1796, https://doi.org/10.1038/ncomms2792.
[20] C.J. Maltby, J.P.R. Schofield, S.D. Houghton, I. O'Kelly, M. Vargas-Caballero, K. Deinhardt, M.J. Coldwell, A 5′ UTR GGN repeat controls localisation and translation of a potassium leak channel mRNA through G-quadruplex formation, Nucleic Acids Res. 48 (2020) 9822–9839, https://doi.org/10.1093/nar/gkaa699.
[21] Y. Feng, S. Tao, P. Zhang, F.R. Sperti, G. Liu, X. Cheng, T. Zhang, H. Yu, X. Wang, C. Chen, D. Monchaud, W. Zhang, Epigenomic features of DNA G-quadruplexes and their roles in regulating rice gene transcription, Plant Physiol. 188 (2022) 1632–1648, https://doi.org/10.1093/plphys/kiab566.
[22] A.A. Surani, C. Montiel-Duarte, Native RNA G quadruplex immunoprecipitation (rG4IP) from mammalian cells, STAR Protoc. 3 (2022) 101372, https://doi.org/10.1016/j.xpro.2022.101372.
[23] R. Hänsel-Hertsch, D. Beraldi, S.V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, D. Tannahill, S. Balasubramanian, G-quadruplex structures mark human regulatory chromatin, Nat. Genet. 48 (2016) 1267–1272, https://doi.org/10.1038/ng.3662.
[24] S.Y. Yang, P. Lejault, S. Chevrier, R. Boidot, A.G. Robertson, J.M.Y. Wong,

D. Monchaud, Transcriptome-wide identification of transient RNA G-quadruplexes in human cells, Nat. Commun. 9 (2018) 1–11, https://doi.org/10.1038/s41467-018-07224-8.
[25] S.Y. Yang, D. Monchaud, J.M.Y. Wong, Global mapping of RNA G-quadruplexes (G4-RNAs) using G4RP-seq, Nat. Protoc. (2022) 1–20, https://doi.org/10.1038/s41596-021-00671-6.
[26] K. Zheng, J. Zhang, Y. He, J. Gong, C. Wen, J. Chen, Y. Hao, Y. Zhao, Z. Tan, Detection of genomic G-quadruplexes in living cells using a small artificial protein, Nucleic Acids Res. 48 (2020) 11706–11720, https://doi.org/10.1093/nar/gkaa841.
[27] Y. Feng, Z. He, Z. Luo, F.R. Sperti, I.E. Valverde, W. Zhang, D. Monchaud, Side-by-side comparison of G-quadruplex (G4) capture efficiency of the antibody BG4 versus the small-molecule ligands TASQs, iScience 26 (2023) 106846, https://doi.org/10.1016/j.isci.2023.106846.
[28] I. Renard, M. Grandmougin, A. Roux, S.Y. Yang, P. Lejault, M. Pirrotta, J.M.Y. Wong, D. Monchaud, Small-molecule affinity capture of DNA/RNA quadruplexes and their identification in vitro and in vivo through the G4RP protocol, Nucleic Acids Res. 47 (2019) 5502–5510, https://doi.org/10.1093/nar/gkz215.
[29] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data, 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
[30] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.
[31] A. Frankish, M. Diekhans, I. Jungreis, J. Lagarde, J.E. Loveland, J.M. Mudge, C. Sisu, J.C. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, C. Boix, S. Carbonell Sala, F. Cunningham, T. Di Domenico, S. Donaldson, I.T. Fiddes, C. García Girón, J.M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, K.L. Howe, T. Hunt, O.G. Izuogu, R. Johnson, F.J. Martin, L. Martínez, S. Mohanan, P. Muir, F.C.P. Navarro, A. Parker, B. Pei, F. Pozo, F.C. Riera, M. Ruffier, B.M. Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, M.Y. Wolf, J. Xu, Y.T. Yang, A. Yates, D. Zerbino, Y. Zhang, J.S. Choudhary, M. Gerstein, R. Guigó, T.J.P. Hubbard, M. Kellis, B. Paten, M.L. Tress, P. Flicek, Gencode 2021, Nucleic Acids Res. 49 (2020) D916–D923, https://doi.org/10.1093/nar/gkaa1087.
[32] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21, https://doi.org/10.1093/bioinformatics/bts635.
[33] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics 30 (2014) 923–930, https://doi.org/10.1093/bioinformatics/btt656.
[34] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.
[35] F. Ramírez, D.P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis, Nucleic Acids Res. 44 (2016) W160, https://doi.org/10.1093/nar/gkw257. –W165.
[36] A.R. Quinlan, BEDTools: the Swiss-army tool for genome feature analysis, Curr Protoc. Bioinform. 47 (2014) 11, https://doi.org/10.1002/0471250953.bi1112s47, 12.1-11.12.34.
[37] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, X.S. Liu, Model-based analysis of ChIP-seq (MACS), Genome Biol. 9 (2008) R137, https://doi.org/10.1186/gb-2008-9-9-r137.
[38] J.T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26, https://doi.org/10.1038/nbt.1754.
[39] M.-A. Turcotte, J.-M. Garant, H. Cossette-Roberge, J.-P. Perreault, Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-quadruplex structures in genes associated with Parkinson's disease, RNA Biol. 18 (2021) 1339–1353, https://doi.org/10.1080/15476286.2020.1847866.
[40] J.-D. Beaudoin, J.-P. Perreault, 5′-UTR G-quadruplex structures acting as translational repressors, Nucleic Acids Res. 38 (2010) 7022–7036, https://doi.org/10.1093/nar/gkq557.
[41] J.-L. Mergny, J. Li, L. Lacroix, S. Amrane, J.B. Chaires, Thermal difference spectra: a specific signature for nucleic acid structures, Nucleic Acids Res. 33 (2005) e138, https://doi.org/10.1093/nar/gni134.
[42] M. Allen, M. Bjerke, H. Edlund, S. Nelander, B. Westermark, Origin of the U87MG glioma cell line: good news and bad news, Sci. Transl. Med. 8 (2016), https://doi.org/10.1126/scitranslmed.aaf6853, 354re3-354re3.
[43] A. Piovesan, M.C. Pelleri, F. Antonaros, P. Strippoli, M. Caracausi, L. Vitale, On the length, weight and GC content of the human genome, BMC Res. Notes 12 (2019) 106, https://doi.org/10.1186/s13104-019-4137-z.
[44] S.G. Rouleau, J.-D. Beaudoin, M. Bisaillon, J.-P. Perreault, Small antisense oligonucleotides against G-quadruplexes: specific mRNA translational switches, Nucleic Acids Res. 43 (2015) 595–606, https://doi.org/10.1093/nar/gku1311.
[45] M.J. Morris, Y. Negishi, C. Pazsint, J.D. Schonhoft, S. Basu, An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES, J. Am. Chem. Soc. 132 (2010) 17831–17839, https://doi.org/10.1021/ja106287x.
[46] S. Lattmann, M.B. Stadler, J.P. Vaughn, S.A. Akman, Y. Nagamine, The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme, Nucleic Acids Res. 39 (2011) 9390–9404, https://doi.org/10.1093/nar/gkr630.
[47] J.-M. Garant, J.-P. Perreault, M.S. Scott, G4RNA screener web server: user focused interface for RNA G-quadruplex prediction, Biochimie 151 (2018) 115–118, https://doi.org/10.1016/j.biochi.2018.06.002.
[48] G.R. Bishop, J.B. Chaires, Characterization of DNA structures by circular dichroism, Curr. Protoc. Nucleic Acid Chem. 11 (2002) 7, https://doi.org/10.1002/0471142700.nc0711s11, 11.1-7.11.8.
[49] J.M. Nicoludis, S.P. Barrett, J.-L. Mergny, L.A. Yatsunyk, Interaction of human telomeric DNA with N- methyl mesoporphyrin IX, Nucleic Acids Res. 40 (2012) 5432–5447, https://doi.org/10.1093/nar/gks152.
[50] N. Arnaud-Barbe, V. Cheynet-Sauvion, G. Oriol, B. Mandrand, F. Mallet, Transcription of RNA templates by T7 RNA polymerase, Nucleic Acids Res. 26 (1998) 3550–3554.
[51] R. Del Villar-Guerra, J.O. Trent, J.B. Chaires, G-quadruplex secondary structure obtained from circular dichroism spectroscopy, Angew Chem. Int. Ed. Engl. 57 (2018) 7171–7175, https://doi.org/10.1002/anie.201709184.
[52] P. Koukouraki, E. Doxakis, Constitutive translation of human α-synuclein is mediated by the 5′-untranslated region, Open Biol. 6 (2016) 160022, https://doi.org/10.1098/rsob.160022.
[53] D.S. McAninch, A.M. Heinaman, C.N. Lang, K.R. Moss, G.J. Bassell, M. Rita Mihailescu, T.L. Evans, Fragile X mental retardation protein recognizes a G quadruplex structure within the survival motor neuron domain containing 1 mRNA 5′-UTR, Mol. Biosyst. 13 (2017) 1448–1457, https://doi.org/10.1039/c7mb00070g.
[54] S. Asamitsu, M. Takeuchi, S. Ikenoshita, Y. Imai, H. Kashiwagi, N. Shioda, Perspectives for applying G-quadruplex structures in neurobiology and neuropharmacology, Int. J. Mol. Sci. 20 (2019) 2884, https://doi.org/10.3390/ijms20122884.
[55] J.F. Moruno-Manchon, P. Lejault, Y. Wang, B. McCauley, P. Honarpisheh, D.A. Morales Scheihing, S. Singh, W. Dang, N. Kim, A. Urayama, L. Zhu, D. Monchaud, L.D. McCullough, A.S. Tsvetkov, Small-molecule G-quadruplex stabilizers reveal a novel pathway of autophagy regulation in neurons, Elife 9 (2020) e52283, https://doi.org/10.7554/eLife.52283.
[56] J.S. Amberger, C.A. Bocchini, F. Schiettecatte, A.F. Scott, A. Hamosh, OMIM.org: online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders, Nucleic Acids Res. 43 (2015) D789–D798, https://doi.org/10.1093/nar/gku1205.
[57] J. Zhao, E.Y.-C. Chow, P.Y. Yeung, Q.C. Zhang, T.-F. Chan, C.K. Kwok, Enhanced transcriptome-wide RNA G-quadruplex sequencing for low RNA input samples with rG4-seq 2.0, BMC Biol. 20 (2022) 257, https://doi.org/10.1186/s12915-022-01448-3.